



Metagenomic Detection Methods in Biopreparedness Outbreak Scenarios

Karlsson, Oskar Erik; Hansen, Trine; Knutsson, Rickard; Löfström, Charlotta; Granberg, Fredrik; Berg, Mikael

Published in:
Biosecurity and Bioterrorism

Link to article, DOI:
[10.1089/bsp.2012.0077](https://doi.org/10.1089/bsp.2012.0077)

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Karlsson, O. E., Hansen, T., Knutsson, R., Löfström, C., Granberg, F., & Berg, M. (2013). Metagenomic Detection Methods in Biopreparedness Outbreak Scenarios. *Biosecurity and Bioterrorism*, 11(1), S146-S157. <https://doi.org/10.1089/bsp.2012.0077>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



METAGENOMIC DETECTION METHODS IN BIOPREPAREDNESS OUTBREAK SCENARIOS

Oskar Erik Karlsson, Trine Hansen, Rickard Knutsson, Charlotta Löfström, Fredrik Granberg,
and Mikael Berg

In the field of diagnostic microbiology, rapid molecular methods are critically important for detecting pathogens. With rapid and accurate detection, preventive measures can be put in place early, thereby preventing loss of life and further spread of a disease. From a preparedness perspective, early detection and response are important in order to minimize the consequences. During the past 2 decades, advances in next-generation sequencing (NGS) technology have changed the playing field of molecular methods. Today, it is within reach to completely sequence the total microbiological content of a clinical sample, creating a metagenome, in a single week of laboratory work. As new technologies emerge, their dissemination and capacity building must be facilitated, and criteria for use, as well as guidelines on how to report results, must be established. This article focuses on the use of metagenomics, from sample collection to data analysis and to some extent NGS, for the detection of pathogens, the integration of the technique in outbreak response systems, and the risk-based evaluation of sample processing in routine diagnostics labs. The article covers recent advances in the field, current debate, gaps in research, and future directions. Examples of metagenomic detection, as well as possible applications of the methods, are described in various biopreparedness outbreak scenarios.

THE CONCEPT OF PREPAREDNESS as applied to biological threats is not new. The Science for Peace and Security program adopted ideas related to this theme in 1956.¹ Biological threats directed toward animals have also been described in historical records and, more recently, the use of anthrax was reported as having been used as a weapon during World War I.^{1A} In the field of animal health, the World Animal Health Organization (OIE) produces and distributes manuals, guidelines, and recommendations as well as standardized methods for veterinary science. In addition, the World Health Organization (WHO) has published recommendations on how to prepare for terrorist

threats to food.² Given the nature of pathogens and the widespread consequences of their release, there is a need for a structured international framework and a coherent view in battling outbreaks.

In biopreparedness there are several subsections, all equally important, and they can be divided in 2 groups: capacity building and capacity evaluation.³ Capacity is built according to intelligence and risk assessment; training is then required to evaluate effectiveness and provide further input for capacity building. Capacity build-up and research fulfill important roles as sources of intelligence, tools to battle threats, and a methodology for implementation,

Oskar Erik Karlsson is a PhD student, Fredrik Granberg, PhD, is Research Coordinator, and Mikael Berg is Professor and Head of the Virology Section; all at the Department of Biomedical Sciences and Veterinary Public Health, Swedish University for Agricultural Sciences, Uppsala, Upland, Sweden. Trine Hansen, MSc, is a PhD student, and Charlotta Löfström, PhD, is Assistant Professor; both at the National Food Institute, Technical University of Denmark, Søborg, Denmark. Rickard Knutsson, PhD, is Director of Security Department, National Veterinary Institute (SVA), Uppsala, Sweden.

evaluation, and risk assessment. The AniBioThreat European Union project aims at increasing the European Union's (EU) capacity to counter animal biological threats by building biopreparedness.⁴

ANIMAL BIOLOGICAL THREATS

Animal biological threats can be of 2 types: threats directed toward animals and threats aimed at the human population by distribution through animals.⁵ With direct threats, the production chain is targeted, and such attacks can lead to economic loss as well as having a socioeconomic impact and an effect on morale, in addition to the logistical challenges of handling the outbreak itself. In the second type, animals are a means to an end, the human target. The OIE estimates that up to 80% of the pathogenic agents suggested for use in bioterrorism have zoonotic origins and as such can be harbored by both animal and human hosts.⁶ Given the often close proximity of farm production plants to populated areas, as well as the high density of animals in production facilities, large-scale outbreaks, whether artificial or natural, can have serious economic and health consequences if not controlled.

DETECTION METHODS

Rapid detection and proper risk assessment are crucial tools in countering biological threats. It is important to quickly detect and confirm the specific infection for effective outbreak control. Studies from Canada estimated that a 1-week delay in implementing control measures for a severe acute respiratory syndrome resulted in a 2.6-fold increase in the mean epidemic size and an extension of 4 weeks of the mean epidemic duration.⁷ The traditional culture-based methods, as well as the immunologic and nucleic acid-based methods, suffer from a number of drawbacks when dealing with novel or unknown organisms. There are no validated culture-based methods for all organisms, and sometimes methods have poor specificity. Immunological and nucleic acid-based methods have a high specificity, making them unable to (or to very little extent) detect organisms outside of their target range. In contrast, metagenomics, a method in which the combined genomes of all organisms present in a sample are analyzed, represents a broad range, high-throughput methodology for detection that requires little or no prior knowledge of the target. This can be of crucial importance when encountering a novel or unexpected target, as it removes the initial lag caused by identification issues.

METAGENOMICS

Metagenomics, the genetics subfield studying the combined genomes of a sample and their interaction, is a quickly

growing field. Not only does it offer a rapid high-throughput method for mapping interactions in a microbial community, but it also circumvents the need for culturing, which enables the study of the unculturable majority of microorganisms in a microbial community.⁸ Next-generation sequencing (NGS) has been suggested as a future application that can improve the response to a bioterrorism incident in the food chain.⁹ Metagenomics can, for instance, be useful in enabling earlier discovery of novel emerging viruses and bioterrorism incidents.¹⁰⁻¹² However, using metagenomics approaches in bioterrorism response requires interoperability between biodefense databases.¹³ Metagenomics also represents a new logistics problem for laboratories implementing the technology: data storage. The falling price of sequencing not only allows for large-scale sequencing investigations but also enables production of enormous amounts of data. Given that the price of storage is steadily increasing and that the computational overhead for analysis is rising with dataset complexity, this is one of the largest obstacles when implementing large-scale metagenomics for screening and diagnostics.¹⁴

Target-Specific Approach

The traditional approach for DNA sequencing was introduced in the 1970s by Sanger, and this method is capable of retrieving up to 1 kb of sequence data at a time.¹⁵ Even though Sanger sequencing has been used for several metagenomics projects, the high demand for low-cost sequencing has caused a shift to use of NGS platforms.^{16,17} Because of the high genetic diversity of most microbial communities, it is currently difficult to obtain enough sequence depth to sample any gene with sufficient coverage to get meaningful information on the diversity or population characteristics. To overcome this difficulty, target-specific metagenomics, which aims to obtain sequence reads from specific genes of interest based on prior knowledge of sequence information, can be a useful tool.¹⁸

The first step in a metagenomic investigation is the amplification of single targets using polymerase chain reaction (PCR); the products are then sequenced either by Sanger sequencing or NGS methods.^{19,20} Target-specific metagenomics generates low amounts of data compared with whole-genome sequencing (WGS), thus allowing analysis of pathogen diversity in communities in a shorter time. This method is therefore useful for screening of pathogens in samples before WGS.

Whole-Genome Approach

The whole-genome approach was pioneered to detect viruses using broad range PCR methods in the late 1990s and early 2000s.^{21,22} With the introduction of affordable sequencing, and in the wake of the technological developments during the Human Genome Project, massive

paralleled Sanger sequencing became a reality and with it the ability to sequence large amounts of clonal libraries at once.²³ Following the first few pioneering experiments, several important discoveries were made within 5 years, proving that our understanding of viral diversity in host species (in this case, human) was so far largely unexplored.²⁴⁻²⁷ Given the introduction of NGS platforms in 2005, the field quickly expanded to encompass this new technological breakthrough. Several successful studies and proof of concept articles were produced in a relatively short span, and in 2012, a total of 198 articles using deep sequencing for detection of viruses were published. Whole-genome based approaches also were used for the bioterrorism investigation of the anthrax letters in the Amerithrax case.²⁸

The whole-genome methodology involves the isolation of genetic material from the sample, enrichment toward the target, enrichment using nonspecific amplification strategies, and finally deep sequencing of the sample. Strategies for sample preparation (extraction of genomic material and enrichment of target-specific material) and amplification of material vary greatly between groups.²⁹ Some favor the unmodified version that was first pioneered (with both enrichment toward target organism as well as amplification), while others suggest that this might introduce undue bias into the datasets.³⁰ Both methodological approaches have merit and are proven principles described in several articles in the literature.

TECHNICAL REVIEW OF METAGENOMICS

Two things are of outmost importance for metagenomic applications: to preserve the metagenome as intact as possible, and to provide as much contextual information as possible. For possible forensic or epidemiologic use, the contextual metadata provided with the sample might be just as important as the sample itself.

Sampling

Metagenomics presents a unique challenge to first responders and field personnel performing the sampling when it comes to deciding what to sample. Because metagenomics has a proven efficiency for detection in solids, liquids, and environmental samples, the need for awareness of one's surroundings for correct sampling is essential. There are many different sample types to be considered, including clinical, water, vegetables, feed, and environmental samples. These sample types have various metadata that should be evaluated for complete analysis of the results.

For clinical samples, the OIE terrestrial manual is the gold standard for methodologies considering animal health and animal welfare, including disease control and surveillance sampling.^{31,32} The main samples to be taken from a

live animal are blood, feces, skin, genital tract swabs and semen, eye and nasal discharge, and milk. For postmortem investigations, the normal guidelines for necropsy should be applicable.³³ When sending samples, a risk assessment according to OIE guidelines, as well as the guidelines for the select agents list, should be made concerning possible risks associated with entry, exposure, and the consequences thereof.³⁴

Sample Treatment

Depending on the sample type, different pretreatments can be used. Each sample type requires a specific protocol, and various types of methods for extraction of nucleic acids are available.³⁵⁻³⁷ If extraction of nucleic acids of microorganisms is associated with a host (eg, plants or animal parts), selective lysis are required to ensure minimal contamination of nucleic acids from the host in the subsequent analysis.³⁸

For liquid samples, selective filtration is an opportunity, whereas for solid samples separation and isolation of cells is an alternative strategy.^{16,39,40} Liquid samples can be represented by both water and various fluids from animals. The ability of the metagenomics approach to detect not only known but also unexpected and unknown microorganisms has been extensively evaluated and proven to work for liquid samples.^{24,41-43} The general procedure is to concentrate the sample and then take a fraction of the sample for sequencing. Solid material can be from environmental samples (eg, swabs from surfaces), food, feed, or whole organs or partial animal parts. Extraction from soil and sediments (eg, environmental samples) is often more difficult than pure cultures because of the presence of inhibitors. The inhibitor can originate from soil compounds that might inhibit the extraction or compounds that copurify with the nucleic acids, such as humic acids, which can interfere with downstream applications.⁴⁴⁻⁴⁶ For these types of samples more thorough sample treatments are needed.⁴⁷ The sample preparation has been shown to influence results from metagenomics analysis—for example, when introducing a freezing step to store the samples before analysis⁴⁸ and when starting with a culture enrichment step.⁴⁹

Since 9/11 and the Amerithrax case, microbial forensics has been developing at great speed. When a forensic investigation of an outbreak is performed, it would be prudent to perform the examination in the way that human medical forensics are conducted, if results are to be included in a possible forensic investigation at a later point. The literature lists several procedures to be taken for preserving and stabilizing samples used for forensic microbiology.⁵⁰ Three main areas can be of interest: extraction techniques and applications, preserving and stabilizing samples, and assessing and controlling sample contamination. All of these are extensively covered in other literature and will not be

discussed here in more detail.³² Figure 1 provides a schematic view of sampling and sample preparation procedures.

Sequencing Technologies

For a considerable amount of time, the long read length paradigm held true for metagenomic studies; thus, technology was chosen based on how long a read it could

produce. This resulted in a number of high-profile novel virus articles, as well as microbial 16S rDNA sequencing to characterize microbial communities. As read length was the main criterion for choosing technology, the Massively Parallel Pyrosequencing, 454 Roche, was the primary choice in metagenomic studies incorporating NGS.^{39,51-53}

The introduction of the Four-Color Reversible Chain Termination (Illumina, HiSeq) technology allowed

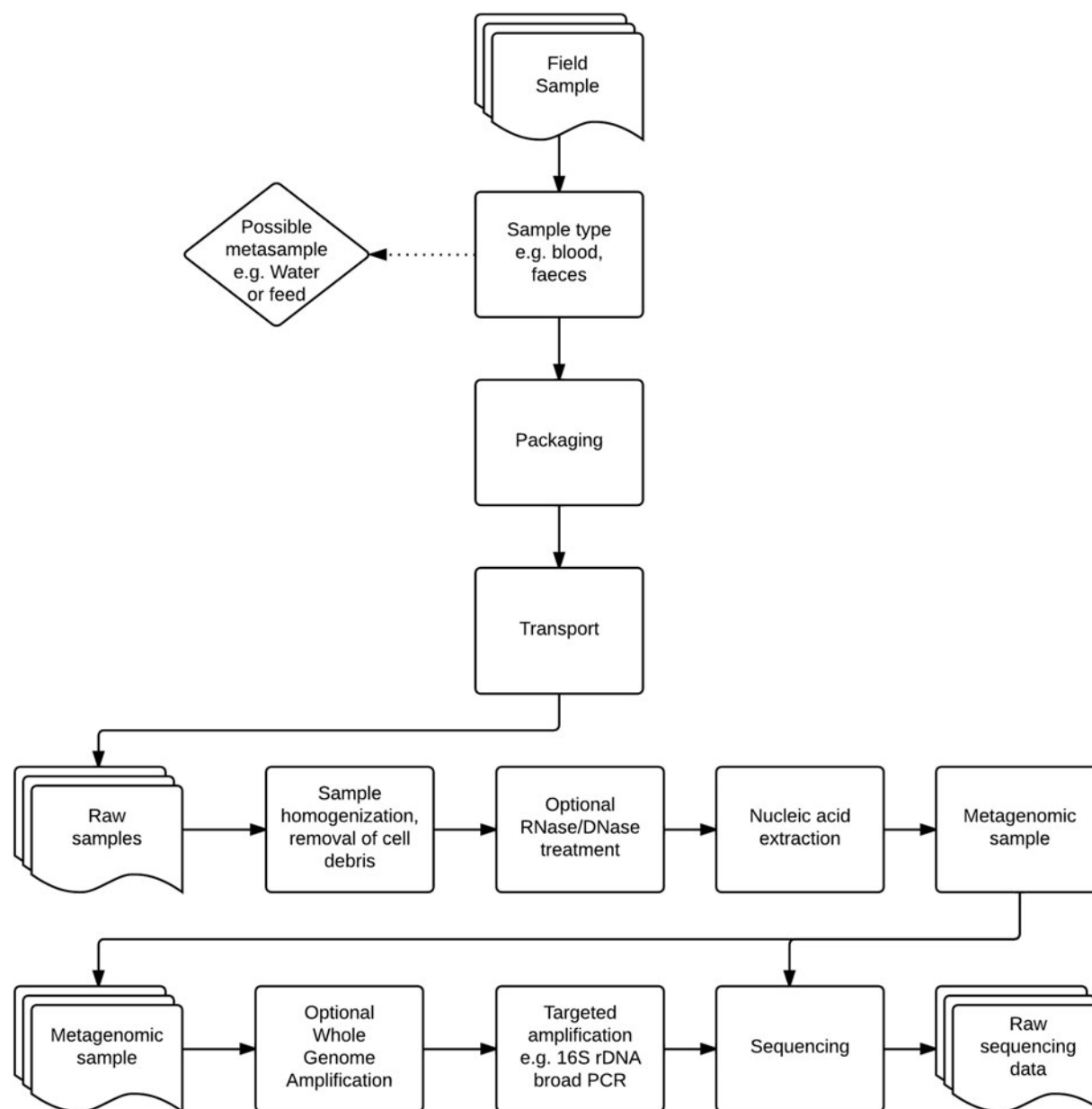


Figure 1. Samples are collected in the field. Following classification and collection of possible metasamples (samples related to the incident site such as feed or water for the animals), risk assessment is performed on site, and samples are packaged in accordance with OIE recommendations and international rules for transport of biological specimens. As samples arrive at the diagnostic lab, they are homogenized, cell debris is removed, and, depending on agents of interest, an optional DNase/RNase step is performed to degrade parts of the metagenome. Nucleic acid extraction is performed on the samples, and they are stored. Samples can now pass through 2 optional steps: whole genome amplification or targeted amplification (eg, 16S rDNA amplification). Sequencing is performed, and raw sequencing data are produced.

unbiased broad-range sequencing of the microbial community. In this approach, relatively short reads are sorted and classed into subgroups and thereafter analyzed. Even though this method has proven useful, the sheer amount of sequence reads, as well as time-consuming analysis and lack of specificity, are factors limiting an extensive use of this technique.⁵⁴⁻⁵⁷

In 2011 LifeTechnology released the Personal Genome Machine (PGM). The PGM employs a technique called sequential electrical detection, in which sequencing of the template strand is performed inside a microchip, directly transferring the released ions into the semiconductor, thereby reading the sequence by the current produced.^{58,59}

Choosing Sequencing Technologies

The choice of sequencing technology for metagenomic investigations is based on 4 main factors: length of sequences, depth of sequencing/throughput of technology, speed of sequencing, and automation availability.⁶⁰⁻⁶² Length of sequences still plays a major role in allowing easy alignment toward genomes of interest, spanning possible repetitive regions, and providing sequence islands in the target genome for further sequencing.⁶³ Depth and

throughput of sequencing depends on (1) how selective you have to be with your sample (selection of parts of the microbiome or presequencing sample preparation), and (2) how many samples you can multiplex (design of assay). The speed of sequencing is of utmost importance from a biopreparedness perspective; technologies taking up to 2 weeks to finish a run are of limited use in an outbreak situation. Finally, automation and the possibility of making the metagenomic approach interoperable in the current automated diagnostic pipeline are factors to consider.

Table 1 presents a compilation of data as given by the sequencer manufacturers adjusted for experimental results, including possible applications, costs for running and acquiring the instrument, and speed for a single run. In 2012, Loman et al provided an excellent initial overview of the newly introduced benchtop sequencers.⁶⁰ The IonProton was introduced after the experiment, as were the increased read-length chemistry for MiSeq and IonTorrent. Suggesting a single technique is therefore complicated. At the time of the writing of this article, MiSeq held the lead in throughput for the 250bp chemistry and has a compelling sample preparation, whereas the IonTorrent provides longer read lengths and faster sequencing turnover. The capabilities of the benchtop sequencers do, however, regardless

Table 1. List of Sequencing Technologies, Their Properties, and Possible Applications

<i>Technology</i>	<i>Applications</i>	<i>Minimum Throughput (Read Length)</i>	<i>Run Time</i>	<i>Cost (Run Price)</i>
454 (Roche)	Small genome, targeted sequencing, transcriptome, metagenomics	700 Mb (up to 1kb)	23 hours	~ \$500k (~ \$6k)
454 Junior	Small genome, targeted sequencing	35 Mb (~ 400b)	10 hours	\$125k (~ \$1k)
Illumina HiSeq	Eukaryote whole genome, exome, targeted sequencing, small genome, transcriptome, metagenomics	600/120 Gb (2 × 100/150)	11 days/ 27 hours	\$690k/\$740k (~ \$23k)
Illumina MiSeq	Exome, small genome, targeted sequencing	~ 4 Gb (2 × 150)	24 hours	\$125k (\$965)
SOLiD	Whole eukaryote genome, exome, transcriptome, metagenomics, small genome, targeted sequencing	95 Gb (2 × 60)	6 days	\$595k (~ \$10k)
SOLID Wildfire	Whole eukaryote genome, exome, transcriptome, metagenomics, small genome, targeted sequencing	240 Gb (2 × 50)	10 days	\$70k upgrade (~ \$5k)
Ion Torrent 314	Small genome, targeted sequencing, metagenomics	10-40Mb (200b)	2-3 hours	\$50k (\$350)
Ion Torrent 316	Small genome, targeted sequencing, metagenomics	100-400Mb (200b)	2-3 hours	\$50k (\$550)
Ion Torrent 318	Small genome, targeted sequencing, metagenomics	~ 1Gb (200b)	2-3 hours	\$50k (\$750)
Ion Proton I	Eukaryote whole genome, exome, targeted sequencing, small genome, transcriptome, metagenomics	~ 10Gb (100b)	2-4 hours	\$149k (\$10k)
Ion Proton II	Eukaryote whole genome, exome, targeted sequencing, small genome, transcriptome, metagenomics	~ 30Gb (100b)	2-4 hours	\$149k (~ \$10k)

Note. The list includes the benchtop version if possible (454 Junior, Illumina MiSeq, and Ion Torrent/Proton Torrent), speed of sequencing as well as possible mean read length, and total data output together with procurement costs and running costs.

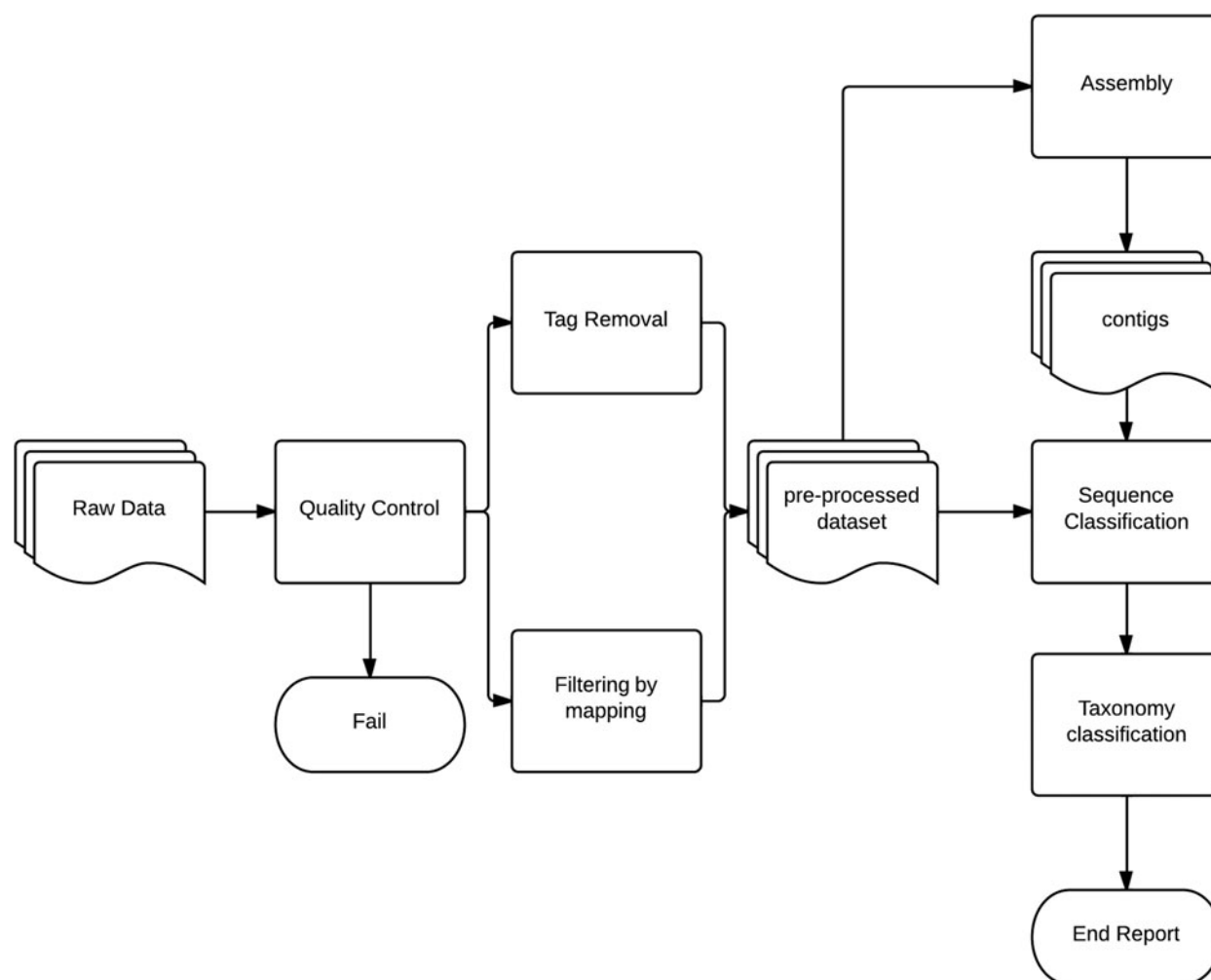


Figure 2. Raw data from sequencing is checked for quality irregularities. Tags from sequencing technology and/or preamplification step are removed, and reads are optionally filtered toward host genome or known contaminating genome (eg, process controls, etc). The preprocessed dataset is split into 2 paths, 1 performing assembly of the reads and 1 mapping the reads directly. Products from assembly are then subjected to the same sort of homology search as the single reads. Finally, the datasets are combined and evaluated for taxonomic classification.

of platform, provide medium-sized laboratories with a good capability to provide rapid metagenomic analysis of clinical samples on a medium scale. The technology will also likely be developed considerably in the near future, making it simpler and having even more capacity.

Bioinformatics and Processing of Data

An integral part of NGS technologies is the postsequence processing of data. After the data have been produced by the chosen sequencing technology, the output is often an abundance of sequence reads with little or no meaning to the untrained eye. Postsequencing processing is applied to raw sequence data to perform quality control (QC), filter reads for contamination, map reads toward relevant genomes, assemble genomes (or partial genomes), and perform homology searches and taxonomic/phylogenetic studies.^{29,64,65} This is

shown in Figure 2 as a flowchart of methodologies and implementation of methodologies for analysis.

Quality of datasets for metagenomic analysis consists of several parts. First, base calling is generally performed by the vendor implementation, crafted into the technology itself, filtering away reads of dubious quality and providing the user with an output from the sequencer.⁶⁶ After the initial filtering, the resulting dataset is usually composed in a data format containing all the relevant information in 1 file. Most vendor pipelines also include the possibility for directly mapping the data toward a genome, giving the user the possibility to map sequences toward several microbial organisms or to map sequences toward the host genome for removal. Mapping of reads toward a genome is functional when a known genome is present; this can either be a host genome (to more efficiently filter the reads and get an estimate of host contamination) or added process controls.^{67,68} This step is normally not used extensively for metagenomic analysis since the strength lies not

on what you know is in the sample but in the broad range of detection. Users can then apply several different quality checks on the dataset to determine its validity for the experiment.

During processing the identification and removal of systematic artifacts is also of great importance, as bias introduced this way can skew the analysis and provide a shifted final picture of the metagenome.⁶⁹ By combining several tools, the user can remove most artificial duplicate reads, clean away tag sequences, assess guanine-cytosine content, trim the sequences, and provide initial filtering by quality, ambiguous base, and sequence complexity.⁷⁰ After initial filtering, as much as 80% to 95% of the initial read counts can be removed, depending on the presequencing processing.^{55,71} This provides the user with a considerably easier dataset to handle. Combined with classification of sequence reads, assembly of short reads can render additional information about the organisms present in the metagenome. Because of the complex nature of metagenomic datasets, assembly can introduce various *in silico* artifacts into the dataset.⁷² These artifacts can be of a different nature, but usually include chimerical contigs and ultradeep contigs. Assemblers are continuously developing new applications, and several metagenomic assemblers have been introduced in the past few years.^{73,74} These assemblers aim at taking the complexity of the metagenome into the assembly, thereby offsetting the possible assembly bias. Moreover, these provide individual unique sequences into the final dataset, thus preserving data for the final analysis.

The final step of a metagenomic analysis is the classification of the sequences. Classification is usually performed by homology-based methods, but both composition-based methods and phylogenetic methodologies exist. These 3 main approaches are sometimes also combined for increased speed and accuracy. (See the review by Bazinet and Cummings for extensive information about classification.⁷⁵) It is noteworthy that homology-based methods have high accuracy and can be fast when using a small database for the query. This provides a possible use for rapidly mapping of a metagenomic sample toward, for example, the select agent, giving fast information about the agents present in a sample.⁷⁶

When reporting metagenomic results, it is important to take into consideration the metadata. The Genomic Standards Consortium provides guidelines for reporting metagenomes. Combined with proper sample documentation according to the OIE sampling standards, these guidelines will suffice for most in-depth analyses of metagenomes for several applications, including preparedness questions or basic science.⁷⁷

Metagenomics-Based Approaches in Outbreak Investigations

There are several examples of how metagenomics-based approaches have been used in investigations aimed at

finding the source and origin of a disease outbreak, including bacteria, viruses, and parasites.^{42,78-81} We present 2 examples to explain the usefulness of applying metagenomics in an outbreak situation when other diagnostic tools were found to be insufficient.

Regions in North Rhine-Westphalia, Germany, and in the Netherlands started in 2011 to see a decline in milk production, and several clinical signs such as fever and diarrhea in cows were identified. Conventional methods failed to establish the source of the infection, and reasons for the decreased milk production were unclear. The Friedrich-Loeffler Institute, Germany, therefore launched a metagenomic study using the 454 Genome Sequencer FLX technology. During sequence analysis, a homology toward 7 orthobunyavirus sequences was discovered. Using repeated sequencing of the original library rendering the reads, a partial viral genome was recovered. Filling the gaps with conventional Sanger sequencing, the Schmallenberg virus was described on a genome level.^{43,82}

Another case was the enterohemorrhagic *Escherichia coli* (EHEC) outbreak in Germany in 2011. This case demonstrated how sequencing technologies could be used in the early stages of an outbreak to characterize the causative agent of the outbreak.⁸³ The outbreak strain was shown to be a hybrid, which means that the current molecular methods used for detection, relying on either detection of virulence genes that are linked to known bacterial serotypes or detection of only the virulence genes, were found to be unable to detect and correctly characterize the outbreak strain. The use of WGS enabled identification of sequences that were specific to the outbreak strain.⁷⁸

DISCUSSION

The evolution of high-throughput sequencing (HTS) technologies toward inexpensive equipment and sequencing runs is slowly making an impact in clinical laboratories. As with all technology introductions, laboratories are now faced with the enormous task of standardizing methodologies and procedures and incorporating results from the new technology into the communication and command structure. In outbreak situations with unknown agents, the criteria for detection are, however, somewhat less stringent than in routine diagnostics. Therefore, NGS-based technologies and methodologies can be a powerful tool in preparedness for detecting unexpected microorganisms.⁸⁴

Diagnostic Use of Metagenomics

The major obstacles to diagnostic use of metagenomics are the timeframe of the analysis and the lack of experienced personnel to perform the data analysis. The first point is slowly resolving itself as the benchtop sequencers today are moving into a 2-day preparation and sequence timeframe.

This is comparable to culture-based and serological methods. The second point is more critical, as the lack of trained informatics personnel is at the moment a serious problem. There is to our knowledge no available and thoroughly tested standard NGS methodology for use in an outbreak. Consequently, there is a need to develop standard operating procedures (SOPs) for field-based sampling by veterinarians or first responders, as well as to develop sample preparation protocols in clinical laboratories. Core facilities, for example, in the EU, should engage in interoperability exercises to ensure common understanding of the methodologies, to exchange expertise, and to identify research needs.

In an outbreak situation, where the origin of the causative agent is not known, the extraction of nucleic acids in a representative manner is the major hurdle for success. Extraction of nucleic acids directly from the sample might introduce bias of the purified DNA or RNA, which in the end could produce misleading results. The methods used for lysis of the cells affect the composition of the total nucleic acid present, since harsh lysis methods that are needed for extraction of DNA from some bacteria (eg, Gram-positive bacteria) might degrade DNA or RNA from others.⁸⁵

Current research has been focused on the detection of novel viruses, a process performed differently depending on, for example, group, sample accessibility, experimental planning, and technology used for NGS. In virology, metagenomics is a proven methodology for detecting unexpected or novel viruses in clinical samples.⁸⁴ When it comes to the bacterial field, metagenomics has mainly been used for investigation of large bacterial communities, such as the human gut, or for environmental samples.^{16,86} However, the use for diagnostic purposes is emerging.

Additionally, the discussion about sequence-based microbiology must be taken further. The same caveat as with PCR-based technologies exists; finding DNA or RNA of an organism does not prove the presence of a living or intact form of that organism. Several factors can produce false-positive results, and with the need for some assembly in several NGS technologies the risk for technological artifacts increases. For well over 100 years, Koch's postulates have been seen as a gold standard for detection of pathological entities. Late in the last century, these postulates came under question with the broad introduction of nucleic acid-based methods into diagnostic microbiology. With the subsequent introduction of NGS methods, there is a clear need for microbiologists to redefine the criteria for how to define causative agents, as reviewed extensively by others.^{84,87}

Theoretical Applications in Diagnostic Preparedness

Metagenomics is today one of the few molecular methods for detection of novel microorganisms, as well as a tool with extreme throughput capabilities. It is our belief that in metagenomics there is a methodology for the unbiased

detection of the microflora of a sample, thus offering a diagnostic capability for tracing introduced organisms at an unprecedented level. In other words, metagenomics can detect several pathogenic organisms in 1 sample, including genetically modified versions of known pathogens. Several different disciplines in the preparedness structure could use such a method applied to microbial forensics, smuggling, outbreak investigations, and R&D needs.

Microbial Forensics

Microbial forensics aims at detecting microorganisms from a forensic perspective—that is, detection of microorganisms provides forensic clues to investigators.⁸⁸ In a preparedness structure, the forensic field can be seen as a part of the epidemiologic work of tracking the origin of introduction. There is also the direct forensic approach in the possible deliberate spread of microorganisms where metagenomics can provide both traceability and detection.

Customs and Smuggling

It is feasible to theorize that high-resolution metagenomics will provide profiling indicating the geographic origin of individual animals. As such, one could hypothesize that it is possible to use metagenomic profiles of the gut microbiota to determine the origin of animals. Previously, we discussed the broad range of detection, the ability to handle a large amount of samples, and the multiplexing capability. This fits well into the customs perspective: large amounts of material to be tested, unknown origin, unknown agents, and possible grave consequences if organisms slip the grid. Given the movement of illegal animals in the EU and globally, large-scale sequencing for broad screening might be implemented in the near future. This would also include the legal and illegal trade of food commodities.

Determining the Causative Agent of an Outbreak

In microbiology, and virology in particular, metagenomics is an important technique to strengthen the capability to handle previously uncharacterized microbial agents and to provide diagnostic laboratories with previously unattainable capabilities for handling outbreaks.

Research Needs

In general, the time from sample to proof of presence of a potential pathogen needs to be speeded up considerably. All steps in the analytical chain (Figure 2) can be improved and simplified, as well as automated. At the moment focus should be put on the bioinformatics, to develop easy-to-use programs that ideally can warn of the presence of potential pathogenic microorganisms in a given sample. A follow-up on this would be to, in the same set, suggest primers for fast real-time PCR detection of the detected organisms, coupled to an oligonucleotide synthesizer that synthesizes new

oligonucleotides that can be used for further strengthening of the proof of the presence of the given pathogen.

In virology there is a clear need for validation of methodology toward a larger sample set of domestic animals. Producing a gold standard for metagenomes, artificial or naturally occurring, would also prove beneficial. As with most molecular techniques involving virology, isolation of the complete viral genome from a clinical sample should be standardized and implemented in diagnostic laboratory procedures. As discussed, the need for establishing whole-genome amplification methods, selective toward microorganisms, and incorporating them into the workflow might be of great use, but the increasingly powerful sequencing equipment might evolve to provide solutions that make this need redundant.

Discussion about the cut-off and reporting of results must also be taken into consideration for implementation in a preparedness setting. Possibly harmful information such as the presence of a pathogen on the restricted list or novel organisms must have standardized procedures for validation.

For large-scale viral screening, the metagenomic approach can be combined with bioinformatics solutions to select for only viral genomes, greatly speeding up the process of aligning sequence results toward possible agents. This combined with a backlog procedure for analyzing the whole metagenome, preferably automated, could greatly improve the rate of detection of novel and emerging viruses.

Further research is also needed for bacteria. The major need is for optimized and standardized methods for sampling and extraction of nucleic acids. Standardized methods should also include ways to quantify possible bacterial contamination to investigate if what is seen is out of the ordinary.

Another area that needs more focus is the link between genotype and phenotype. The uncertainty in predicting the phenotype solely on genomic or metagenomic data implies that the culture-based techniques never will be completely replaced.⁸⁹ Furthermore, for certain toxin-producing bacteria (eg, *Clostridium botulinum*) it is not necessarily the bacterium itself that is the causative agent, but the toxins. The presence of active toxins cannot be identified by genome-sequencing techniques. For this reason, other techniques are needed such as the Endopep-MS, in which the presence and activity of toxins is investigated.⁹⁰

CONCLUSIONS

This article has identified potential applications of metagenomics technologies for biopreparedness applications. For instance, metagenomic approaches for molecular detection of microbial agents are important in a biopreparedness perspective and hold great merit for laboratory response capabilities. Not only does it improve on the current molecular methods in throughput and flexibility, it also presents the only currently unbiased high-throughput methodology for broad screening of samples. Metagenomes

as a research field also provides new insights into biomarkers produced by the microorganisms connected to animals, enabling new tracing of food and feed as well as possible forensic applications.

ACKNOWLEDGMENTS

Writing of this publication has been supported by the framework of the EU project AniBioThreat (Grant Agreement: Home/2009/ISEC/AG/191) with financial support from the Prevention of and Fight against Crime Programme of the European Union, European Commission—Directorate General Home Affairs. This publication reflects views only of the authors, and the European Commission cannot be held responsible for any use that may be made of the information contained therein.

REFERENCES

- Science for Peace and Security. Historical context. North Atlantic Treaty Organization website. Updated May 31, 2011. http://www.nato.int/science/about_sps/historical.htm. Accessed July 15, 2013.
- Casagrande R, Willis N. Bioterrorism targeted at agriculture. In: Katz R, Zilinskas RA, eds. *Encyclopedia of Bioterrorism Defense*. 2nd ed. New York: Wiley; 2011:130-137.
- Terrorist Threats to Food – Guidelines for Establishing and Strengthening Prevention and Response Systems*. Geneva: World Health Organization; 2008.
- Hémond Y, Robert B. Preparedness: the state of the art and future prospects. *Disaster Prev Manag* 2012;21(4):404-417.
- AniBioThreat website. 2012. <http://www.anibiothreat.com>. Accessed July 15, 2013.
- Yeh JY, Park JY, Cho YS, Cho IS. Animal biowarfare research: historical perspective and potential future attacks. *Zoonoses Public Health* 2012 Dec;59(8):536-544.
- Ryan CP. Zoonoses likely to be used in bioterrorism. *Public Health Rep* 2008 May-Jun;123(3):276-281.
- Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004 Sep 15;160(6):509-516.
- Pachter L. Interpreting the unculturable majority. *Nat Methods* 2007 Jun;4(6):479-480.
- Knutsson R, van Rotterdam B, Fach P, et al. Accidental and deliberate microbiological contamination in the feed and food chains—how biotraceability may improve the response to bioterrorism. *Int J Food Microbiol* 2011 Mar 1;145(Suppl 1):S123-S128.
- Rosen GL, Sokhansanj BA, Polikar R, et al. Signal processing for metagenomics: extracting information from the soup. *Curr Genomics* 2009 Nov;10(7):493-510.
- Tang P, Chiu C. Metagenomics for the discovery of novel human viruses. *Future Microbiol* 2010 Feb;5(2):177-189.
- Nakamura S, Yang CS, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 2009;4(1):e4219.

13. Valdivia-Granda WA. Bioinformatics for biodefense: challenges and opportunities. *Biosecur Bioterror* 2010 Mar;8(1): 69-77.
14. Desai N, Antonopoulos D, Gilbert JA, Glass EM, Meyer F. From genomics to metagenomics. *Curr Opin Biotechnol* 2012 Feb;23(1):72-76.
15. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977 Dec;74(12):5463-5467.
16. Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004 Apr 2;304(5667):66-74.
17. Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science* 2006 Jun 2;312(5778):1355-1359.
18. Suenaga H. Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* 2012 Jan;14(1): 13-22.
19. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science* 2005 Jun 10;308(5728):1635-1638.
20. Kalyuzhnaya MG, Lapidus A, Ivanova N, et al. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* 2008 Sep; 26(9):1029-1034.
21. Breitbart M, Salamon P, Andresen B, et al. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 2002 Oct 29;99(22):14250-14255.
22. Breitbart M, Hewson I, Felts B, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003 Oct;185(20):6220-6223.
23. Cox-Foster DL, Conlan S, Holmes EC, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 2007 Oct 12;318(5848):283-287.
24. Blomstrom AL, Widen F, Hammer AS, Belak S, Berg M. Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome by use of viral metagenomics. *J Clin Microbiol* 2010 Dec;48(12):4392-4396.
25. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog* 2008 Feb;4(2): e1000011.
26. Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol* 2009 Nov;11(11):2806-2820.
27. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL. Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog* 2008;4(9):e1000163.
28. Rasko DA, Worsham PL, Abshire TG, et al. Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. *Proc Natl Acad Sci U S A* 2011 Mar 22;108(12):5027-5032.
29. Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol* 2005 Jun;3(6):504-510.
30. Kim KH, Bae JW. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl Environ Microbiol* 2011 Nov;77(21): 7663-7668.
31. OIE. *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2013*. World Organisation for Animal Health; 2012. <http://www.oie.int/international-standard-setting/terrestrial-manual/access-online/>. Accessed July 15, 2013.
32. Bayona JM, Lord HL. *Comprehensive Sampling and Sample Preparation: Analytical Techniques for Scientists, Vol. 4*. New York: Elsevier; 2012.
33. Straffuss AC. *Necropsy: Procedures and Basic Diagnostic Methods for Practicing Veterinarians*. Springfield, IL: Charles C Thomas; 1988.
34. Collection, submission and storage of diagnostic specimens. In: *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2013*. World Organisation for Animal Health; 2012.
35. Roh C, Villatte F, Kim BG, Schmid RD. Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl Biochem Biotechnol* 2006 Aug;134(2):97-112.
36. Smith K, Diggle MA, Clarke SC. Comparison of commercial DNA extraction kits for extraction of bacterial genomic DNA from whole-blood samples. *J Clin Microbiol* 2003 Jun;41(6):2440-2443.
37. Kok T, Wati S, Bayly B, Devonshire-Gill D, Higgins G. Comparison of six nucleic acid extraction methods for detection of viral DNA or RNA sequences in four different non-serum specimen types. *J Clin Virol* 2000 Feb;16(1): 59-63.
38. Burke C, Kjelleberg S, Thomas T. Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl Environ Microbiol* 2009 Jan;75(1):252-256.
39. Angly FE, Felts B, Breitbart M, et al. The marine viromes of four oceanic regions. *PLoS Biol* 2006 Nov;4(11):e368.
40. Thomas T, Gilbert J, Meyer F. Metagenomics—a guide from sampling to data analysis. *Microb Inform Exp* 2012; 2(1):3.
41. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U S A* 2005 Sep 6;102(36):12891-12896.
42. Yu G, Greninger AL, Isa P, et al. Discovery of a novel polyomavirus in acute diarrheal samples from children. *PLoS One* 2012;7(11):e49449.
43. Hoffmann B, Scheuch M, Hoper D, et al. Novel orthobunyavirus in Cattle, Europe, 2011. *Emerg Infect Dis* 2012 Mar;18(3):469-472.
44. Sutlovic D, Gamulin S, Definis-Gojanovic M, Gugic D, Andjelinovic S. Interaction of humic acids with human DNA: proposed mechanisms and kinetics. *Electrophoresis* 2008 Apr;29(7):1467-1472.
45. Braid MD, Daniels LM, Kitts CL. Removal of PCR inhibitors from soil DNA by chemical flocculation. *J Microbiol Methods* 2003 Mar;52(3):389-393.
46. Desai C, Madamwar D. Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. *Bioresour Technol* 2007 Mar;98(4): 761-768.
47. Delmont TO, Robe P, Clark I, Simonet P, Vogel TM. Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J Microbiol Methods* 2011 Sep;86(3): 397-400.
48. Bahl MI, Bergstrom A, Licht TR. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes

- ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol Lett* 2012 Apr;329(2):193-197.
49. Pettengill JB, McAvoy E, White JR, Allard M, Brown E, Ottesen A. Using metagenomic analyses to estimate the consequences of enrichment bias for pathogen detection. *BMC Res Notes* 2012;5:378.
50. Budowle B, Schutze SE, Breeze RG, Keim PS, Morse SA, eds. *Microbial Forensics*. 2d ed. Burlington, MA: Elsevier/Academic Press; 2011.
51. Poinar HN, Schwarz C, Qi J, et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 2006 Jan 20;311(5759):392-394.
52. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005 Sep 15;437(7057):376-380.
53. Edwards RA, Rodriguez-Brito B, Wegley L, et al. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 2006;7:57.
54. Bennett ST, Barnes C, Cox A, Davies L, Brown C. Toward the 1,000 dollars human genome. *Pharmacogenomics* 2005 Jun;6(4):373-382.
55. Kostic AD, Ojesina AI, Pedamallu CS, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 2011 May;29(5):393-396.
56. Lazarevic V, Whieson K, Huse S, et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* 2009 Dec;79(3):266-271.
57. Rodrigue S, Materna AC, Timberlake SC, et al. Unlocking short read sequencing for metagenomics. *PLoS One* 2010; 5(7):e11840.
58. Whiteley AS, Jenkins S, Waite I, et al. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* 2012 Oct;91(1):80-88.
59. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011 Jul 21;475(7356):348-352.
60. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012 May;30(5):434-439.
61. Dunne WM Jr., Westblade LF, Ford B. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur J Clin Microbiol Infect Dis* 2012 Aug;31(8):1719-1726.
62. Koser CU, Ellington MJ, Cartwright EJ, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* 2012 Aug;8(8):e1002824.
63. Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. *Appl Environ Microbiol* 2008 Mar;74(5): 1453-1463.
64. Scholz MB, Lo CC, Chain PS. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol* 2012 Feb;23(1): 9-15.
65. Lysholm F, Wetterbom A, Lindau C, et al. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS One* 2012;7(2):e30875.
66. Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. *Brief Bioinform* 2011 Sep;12(5): 489-497.
67. Sachsenroder J, Twardziok S, Hammerl JA, et al. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One* 2012;7(4):e34631.
68. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 2010 Sep; 11(5):473-483.
69. Gomez-Alvarez V, Teal TK, Schmidt TM. Systematic artifacts in metagenomes from complex microbial communities. *ISME J* 2009 Nov;3(11):1314-7.
70. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011 Mar 15;27(6): 863-864.
71. Cheval J, Sauvage V, Frangeul L, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 2011 Sep; 49(9):3268-3275.
72. Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* 2012 Apr;6(4): 898-901.
73. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* 2012 Nov 1;40(20):e155.
74. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 2012 Dec 22;13(12):R122.
75. Bazinet AL, Cummings MP. A comparative evaluation of sequence classification programs. *BMC Bioinformatics* 2012; 13:92.
76. Roux S, Faubladier M, Mahul A, et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics* 2011 Nov 1; 27(21):3074-3075.
77. Field D, Amaral-Zettler L, Cochrane G, et al. The Genomic Standards Consortium. *PLoS Biol* 2011 Jun;9(6): e1001088.
78. Mellmann A, Harmsen D, Cummings CA, et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 2011;6(7):e22751.
79. McMullan LK, Frace M, Sammons SA, et al. Using next generation sequencing to identify yellow fever virus in Uganda. *Virology* 2012 Jan 5;422(1):1-5.
80. Nakamura S, Maeda N, Miron IM, et al. Metagenomic diagnosis of bacterial infections. *Emerg Infect Dis* 2008 Nov;14(11):1784-1786.
81. Kawai T, Sekizuka T, Yahata Y, et al. Identification of *Kudoa septempunctata* as the causative agent of novel food poisoning outbreaks in Japan by consumption of *Paralichthys olivaceus* in raw fish. *Clin Infect Dis* 2012 Apr;54(8):1046-1052.
82. Beer M, Conraths FJ, van der Poel WH. 'Schmallenberg virus'—a novel orthobunyavirus emerging in Europe. *Epidemiol Infect* 2013 Jan;141(1):1-8.
83. Frank C, Faber MS, Askar M, et al. Large and ongoing outbreak of haemolytic uraemic syndrome, Germany, May 2011. *Euro Surveill* 2011;16(21).
84. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012 Feb; 2(1):63-77.

85. von Wintzingerode F, Gobel UB, Stackebrandt E. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 1997 Nov;21(3):213-229.
86. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010 Mar 4;464(7285):59-65.
87. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* 1996 Jan;9(1):18-33.
88. Tucker JB, Koblenz GD. The four faces of microbial forensics. *Biosecur Bioterror* 2009 Dec;7(4):389-397.
89. Pallen MJ, Loman NJ, Penn CW. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. *Curr Opin Microbiol* 2010 Oct;13(5):625-631.
90. Kalb SR, Moura H, Boyer AE, McWilliams LG, Pirkle JL, Barr JR. The use of Endopep-MS for the detection of botulinum toxins A, B, E, and F in serum and stool samples. *Anal Biochem* 2006 Apr 1;351(1):84-92.

*Manuscript received December 20, 2012;
accepted for publication May 17, 2013.*

Address correspondence to:

Oskar Erik Karlsson

Swedish University of Agricultural Sciences

Department of Biomedical Sciences and Veterinary Public Health

Section of Virology Box 7028

Uppsala, Uppland 750 07 Sweden

E-mail: oskar.e.karlsson@slu.se